AI Ethics · Essay

# What Store Managers Taught Me About Automating Hiring Decisions

*A reflection on leading AI governance for a hiring system used across a large multi-brand retail footprint, and what consent, bias, and dignity look like from the applicant's side of the screen.*

Jennifer McKinney · June 10, 2026 · 6 min read

I learned more about AI governance from store managers than from conference panels.

The program was a large conversational hiring-AI deployment for a national, multi-brand retailer, spanning thousands of store locations across multiple countries. The work involved a detailed risk register, CCPA and adverse-action controls, and the predictable mass of delivery work that accompanies any enterprise program. Requirements. Vendor coordination. Integration dependencies. Training. Exceptions. The things that determine whether a well-intentioned design survives contact with a real organization.

But the lessons that stayed with me were human ones.

At headquarters, automation often arrives as a simple proposition. The hourly hiring process is too slow. Applicants drop out. Managers are overwhelmed. A conversational interface can screen for basic qualifications, schedule the next step, and return time to the people who actually have to run the store. All of that can be true.

It can also be incomplete.

For the person on the other end of the conversation, this is not an efficiency initiative. It is a job application. It may be a rent payment, a first job, a return to work after illness, a path out of an unsafe situation, or a second chance after a rough year. The applicant does not experience the system as a workflow optimization. They experience it as a gatekeeper.

That changed the questions I wanted the program to answer.

The first was consent. Too many teams treat consent as a disclosure that can be placed somewhere in the flow and then checked off. I think of it more practically. Does the person understand that they are interacting with an automated system? Do they understand what information is being used? Do they have a meaningful way to ask for help, correct an error, or

choose a different path when the system does not fit their situation?

The distinction matters. A candidate who does not complete a text-based interaction may be uninterested. They may also have limited connectivity, a disability that makes the experience difficult, a language mismatch, or a reasonable distrust of disclosing sensitive information to a chatbot. An efficient system can mistake friction for disinterest. That is not a technical glitch. It is a design choice with consequences.

The second question was bias, but not in the narrow sense of a single model metric. Bias in hiring systems is often discussed as if the only task is to calculate whether one group receives a worse score than another. Measurement matters. It is not enough.

The harder issue is the entire decision path. What questions are asked? Which answers count as evidence? What assumptions are embedded in a definition of “qualified”? What happens when a candidate gives an answer that is unusual but valid? Who notices when an operational shortcut begins to exclude exactly the people a retailer says it wants to hire?

There is no neutral baseline hidden inside a hiring process. The old process also had subjectivity, inconsistency, and delay. That does not make automation wrong. It makes the obligation more serious. When we scale a process, we scale its strengths and its blind spots together.

The managers helped make this obvious. They were closest to the job itself. They knew when a resume did not tell the whole story. They knew that reliability, customer judgment, and the ability to learn could show up in a conversation in ways a rigid screen would miss. They also knew their own time was finite. A governance approach that demanded manual review of everything would have collapsed under the volume. A system that offered no escape from its own decisions would have been irresponsible.

That is where human review thresholds became real rather than rhetorical.

“Human in the loop” is a phrase that can conceal more than it reveals. A human is not meaningfully in the loop if the system has already made the consequential decision and the person is only there to rubber-stamp it. Nor is human review a complete answer if every borderline case is sent to an already overloaded manager without context, policy, or time to act.

We needed specific thresholds and specific paths. High-confidence, low-consequence workflow steps could be automated. Ambiguous cases needed escalation. A candidate facing an adverse outcome needed a legible process, not a black box. A manager needed enough context to make a judgment rather than simply receive a system recommendation. And the program needed evidence that the threshold was functioning as intended, not just a statement that a human could intervene in theory.

The legal dimension was important, but I did not want the program to reduce to legal compliance. A compliance control can specify what the organization must document. It cannot answer, by itself, whether the people affected by the system have been treated with care. For me, that was a useful reminder that technical decisions have labor consequences even when nobody involved intends harm.

The third lesson was about dignity. Dignity is sometimes treated as a soft concept, separate from architecture. I think it is architecture expressed from the applicant's side.

A dignified hiring system says what it is doing. It does not pretend a chatbot is a person. It does not ask for information it cannot justify. It gives applicants a way to correct a misunderstanding. It makes room for a human being to see the person who does not fit the expected pattern. It provides enough explanation that rejection does not feel like a message from an invisible machine.

None of this requires pretending that technology can eliminate the hard parts of hiring. It cannot. Hiring is judgment under uncertainty. The question is whether we make that judgment more accountable or simply less visible.

That is why I came to see the risk register as more than a compliance artifact. Its line items were not interchangeable rows in a spreadsheet. They represented possible failures in trust: an applicant who did not understand the process, a manager who could not override it, an outcome that could not be explained, a data practice that exceeded its purpose, a threshold that turned a useful shortcut into an exclusion mechanism.

Good program management makes those concerns operational. It names the owner. It defines the decision. It records the evidence. It identifies the escalation route. It does not solve every ethical question, but it makes it harder for the organization to evade the question until an incident forces attention.

NIST's Generative AI Profile calls on organizations to incorporate trustworthy characteristics into system requirements proactively. I read that not as a compliance instruction but as a useful discipline. The requirements phase is where a team decides whose inconvenience counts, whose uncertainty triggers review, and what evidence will exist when someone asks how a decision was reached. NIST

The store managers did not ask for an ethics manifesto. They asked for a process that worked, a way to handle exceptions, and confidence that automation would help them see more qualified people rather than hide good candidates behind a screen.

That is still my standard. Automation should remove administrative burden, not remove the person from the decision. It should make hiring more consistent without making it less humane.

And it should be designed with the humility to recognize that every screening system is making a claim about who deserves to be seen.

---

## Sources

- NIST